🌡️ Daily Temperature vs. Water Intake 💧

# Investigating Environmental Factors Influencing Personal Hydration Behavior

## Abstract

Hydration is essential for maintaining core bodily functions such as thermoregulation, metabolism, and cellular health. This study explores the impact of environmental variables—namely temperature and humidity—on daily water consumption. By integrating meteorological data from the Meteostat API with personal hydration logs and recommended intake levels from the WaterMinder application, this project aims to uncover potential behavioral trends and support data-driven hydration recommendations.

## 1. Introduction

While the importance of drinking water is widely acknowledged, many individuals struggle with maintaining adequate hydration, especially when environmental conditions fluctuate. Motivated by a personal observation—that drinking water feels easier on hotter days—this study investigates whether a measurable relationship exists between ambient temperature, atmospheric humidity, and individual water intake.

**1.1 Motivation**

Drinking water is important for everyone’s health. However, for me, it always felt like a task. Since I do sports, I need to pay extra attention to my water intake. But I realized that on hot days, drinking water felt easier. Additionally, I wondered whether temperature and humidity levels also play a role in hydration behavior. This made me curious about the relationship between temperature, humidity, and water consumption, so I decided to analyze it.

The core hypothesis tested in this study is:  
"Individuals are more likely to consume greater amounts of water on hotter days and in drier conditions."

## 2. Data Sources

### 2.1 Temperature & Humidity Data

- Provider: Meteostat API  
- Location: Istanbul Kurtköy/Göztepe (selected due to proximity to daily routine)  
- Format: Excel export of daily average temperature and humidity  
- Reason for Selection: Reliable and easy-to-integrate historical weather data

Note: During data collection via the Meteostat API, missing data was encountered for the Kurtköy station on certain days. To ensure completeness, daily averages were supplemented by combining data from the Göztepe station, which is geographically and climatically comparable.

### 2.2 Water Intake Logs

- Provider: WaterMinder  
- Format: CSV/PDF export of daily water intake in milliliters  
- Reason for Selection: Provides structured and timestamped personal hydration records

### 2.3 Daily Calories Burned by Step

- Provider: Apple watch  
- Format: Excel  
- Purpose: Provides more accurate causation in between data types.

**2.4 Data Integration Notes and Source Clarifications**

During data collection, several files were used and merged to build a unified dataset:

* *‘weather-combined.xlsx’***:** Full weather dataset combining missing entries between *Istanbul Kurtköy* and *Göztepe* weather stations using the Meteostat API. This file ensures complete daily temperature and humidity data.
* ***‘****kurtkoy-weatherp2.xlsx*’ and ‘*weather-goztepe-p2.xlsx’***:** Supplemental files containing extended weather logs from Meteostat, used to fill in missing rows.
* ***‘****wtmndrdata.csv’***:** Water consumption data exported from WaterMinder app, logging daily hydration levels in milliliters.
* **‘***step-calories.xlsx’*: Step count and calorie expenditure data collected from Apple Watch, used to measure daily physical activity.

These files were cleaned, aligned, and merged during the data preparation process. Weather data from Kurtköy had gaps that were filled using Göztepe data to maintain consistency in daily tracking.

## 3. Methodology

### 3.1 Data Preparation

Dates were aligned across datasets. Missing values were handled through row elimination or interpolation (with the help of AI tool). Variables were unified into a single DataFrame: Date, Temperature (°C), Humidity (%), Water Intake (ml), Calories burned by step(cal).

### 3.2 Exploratory Analysis

Visual Tools Used:  
- Scatter plots to visualize relationships between temperature and intake  
- Time-series graphs for daily patterns

- Correlation matrix to see correlations

### 3.3 Correlation Analysis

Statistical Method: Pearson correlation  
Variables Compared:  
- Temperature vs. Water Intake  
- Humidity vs. Water Intake  
- Burned Calories vs. Weather Variables

**3.4 Hypothesis Test: Temperature vs. Water Intake**

To statistically evaluate the hypothesis that individuals drink more water on hotter days, a Pearson correlation-based hypothesis test was conducted. The test evaluated the strength and significance of the relationship between average daily temperature and daily water intake.  
  
Pearson correlation coefficient (r): 0.0873  
One-tailed p-value: 0.2691  
Result: Fail to reject H₀ → No significant positive correlation found.  
  
This result suggests that the observed trend is not statistically significant at the conventional confidence levels and the data does not provide strong evidence to support a positive linear relationship between temperature and daily water intake.

### 3.5 Predictive Modeling

Models Implemented:  
- Random Forest Regressor  
- K-Nearest Neighbors (KNN) Regressor  
Features: Temperature, Humidity, Recommended Intake  
Target: Daily Water Intake  
Outcome: Random Forest showed marginally better accuracy in predicting hydration levels.

## 4. Key Findings

- Positive correlation found between temperature and water intake. (Figure 5)  
- Negative correlation found between humidity and water intake—individuals tend to drink more in drier weather. (Figure 5)  
- Predictive models demonstrate potential for estimating daily intake based on weather. (Figure 5)

## 5. Discussion

The observed correlations support the hypothesis: as temperature increases and humidity drops, individuals are more inclined to drink water, potentially due to physiological mechanisms like increased perspiration. These findings align with environmental behavior theories in health analytics.

## 6. Machine Learning Evaluation

To enhance the predictive insights of the analysis, two machine learning models were implemented: Random Forest Regressor and K-Nearest Neighbors (KNN) Regressor. Both models were used to estimate daily water intake based on temperature, humidity, and calorie values.

### 6.1 Features Usedcs

- Temperature (°C)  
- Humidity (%)  
- Calories Burned(cal)

### 6.2 Model Comparison Table

| Model | Mean Squared Error (MSE) | R² Score (R²) |  
|-----------------------------------|--------------------- |------------|  
| Random Forest Regressor | 27337.10 |-0.78 |  
| K-Nearest Neighbors | 182,488.59 | -0.19 |

### 6.3 Interpretation

While Random Forest is generally expected to outperform KNN in capturing nonlinear relationships, in this dataset the **KNN model performed better overall**, possibly due to its sensitivity to local data structure and less overfitting tendency. However, both models indicate limited predictive power, and further improvements could be achieved through more informative features or a larger dataset.

## 7. Limitations

This study provides an initial exploration of how environmental factors such as temperature and humidity relate to daily water intake behavior. However, there are a number of limitations that should be considered:  
  
- Sample Size: The dataset consists of a limited number of personal observations, which restricts the generalizability of the results.  
- Personal Bias: The data used reflects the behavior of a single individual, limiting the ability to draw population-wide conclusions.  
- External Factors Not Included: Other potentially influential factors such as physical activity duration, diet, or fluid intake from food were not captured.  
- Model Performance:\*Both machine learning models showed weak predictive capabilities, indicating that additional or more relevant features are needed.

## 8. Future Work

Future work may focus on the following:  
- Expanding the dataset to include more participants over a longer duration.  
- Integrating additional features such as sleep, activity type, and nutritional data.  
- Experimenting with advanced modeling techniques and feature engineering.  
- Developing a personalized hydration recommendation system using real-time weather and behavioral data.

## 9. Conclusion

The relationships between variables were explored through visualizations and a correlation matrix . A hypothesis test was conducted to evaluate the correlation between temperature and water intake, but no statistically significant relationship was found.

To extend the analysis, I implemented two machine learning models — Random Forest and K-Nearest Neighbors (KNN) — to predict daily water intake using environmental and physiological features. Interestingly, **the KNN model performed better**, achieving a lower Mean Squared Error and a higher R² score than Random Forest. This suggests that in this dataset, KNN was more effective in capturing the underlying structure, while Random Forest may have overfit to the training data.

**10. Visual Results**

Figure 1:

This graph visualizes daily calories burned based on step data recorded by Apple Watch. It was used to examine whether physical activity levels correlate with water intake. The data was exported in .xlsx format and aligned with daily weather and hydration records.

A graph showing a number of calories burned by step

Description automatically generated

Figure 2:

This chart shows the amount of water consumed daily, which serves as the target variable in both correlation analysis and prediction models. The data was exported from the WaterMinder app and processed to match daily environmental conditions.

A graph with orange lines

Description automatically generated

Figure 3:

This plot illustrates the average temperature recorded each day, used to evaluate its influence on water intake. The data was obtained through the Meteostat API, combining entries from Kurtköy and Göztepe stations to ensure completeness.

A graph showing the temperature of the day

Description automatically generated

Figure 4:

This graph presents daily humidity levels, examined as a potential factor affecting hydration behavior. The data was sourced from the Meteostat API and merged from two nearby stations to fill any missing values.

A graph showing the average humidity

Description automatically generated

Figure 5:

This correlation matrix quantifies the linear relationships between all variables in the dataset, including temperature, humidity, water intake, calories burned, and calories burned by step.It was calculated using the merged and cleaned dataset via Pandas.

A diagram of a graph

Description automatically generated with medium confidence

Figure 6:

This scatter plot depicts the relationship between temperature and daily water intake, with a regression line added to visualize the trend. It was generated using the combined dataset after preprocessing and alignment.

A graph with blue dots and a red line

Description automatically generated

Figure 7:

This figure presents the performance metrics of the Random Forest model, including Mean Squared Error (MSE) and R² score. The model was trained using temperature, humidity, calorie values as predictors.

A graph with green and red dots

Description automatically generated

Figure 8:

This figure summarizes the K-Nearest Neighbors (KNN) model’s performance in predicting water intake. It serves as a comparison point for evaluating different regression models used in the study.

A graph with a red line and purple dots

Description automatically generated